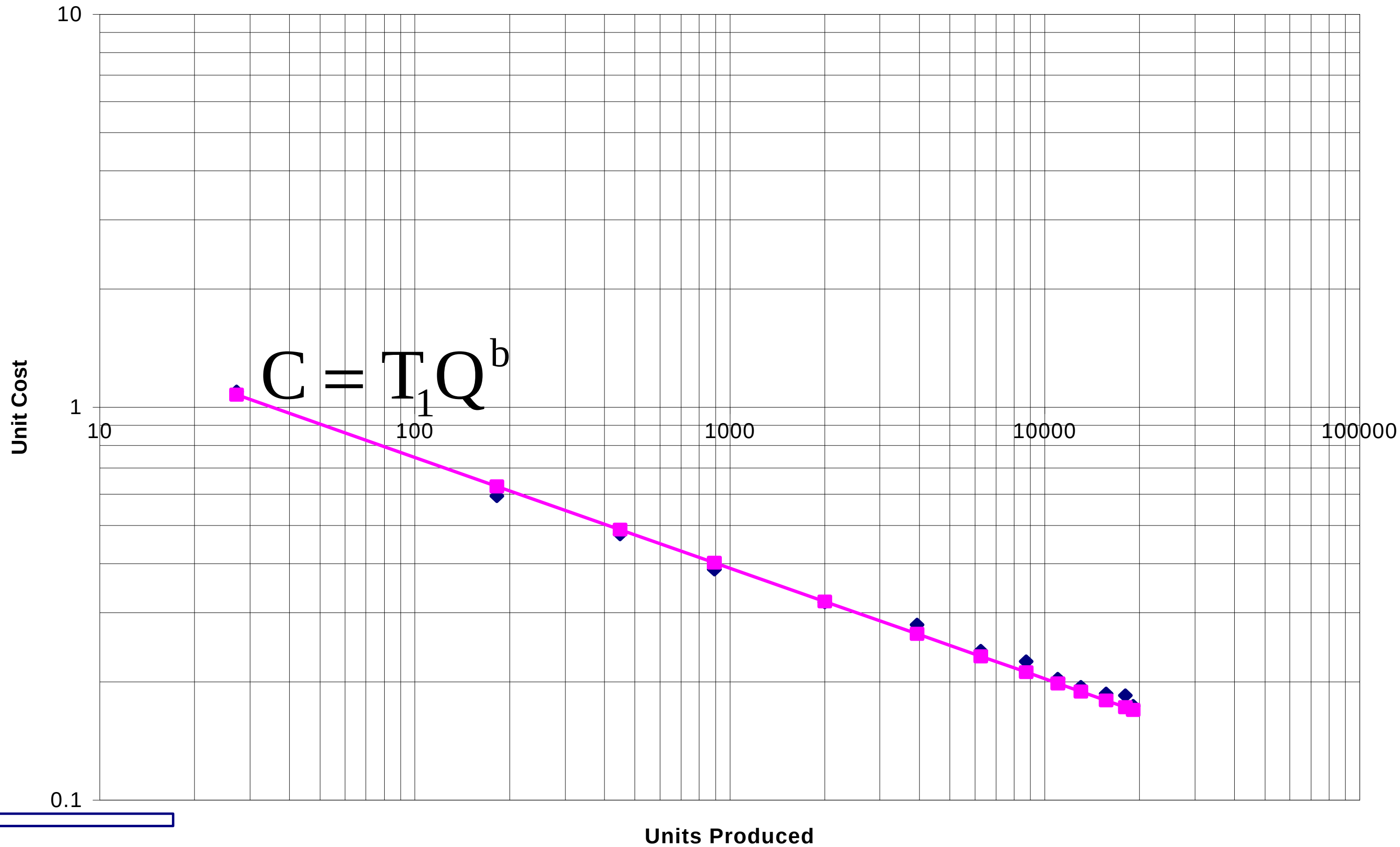
Analyzing Development Programs' Expenditures with the Norden-Rayleigh Model

David Lee
Logistics Management Institute
703-917-7557
dalee@lmi.org

32nd ADODCAS, Wednesday, February 3, 1999

Power-law model of manufacturing cost progress: a great success of parametric analysis



How about development programs?

The Norden/Rayleigh model for time-phasing of expenditures in development programs

- Enjoys strong support from data on actual programs
- Useful for forecasting cost-to-go and time-to-go for development programs, given expenditures vs time for an initial period
- Is a particular case of a more general perspective on development programs' costs

The Rayleigh Model

- Norden (1963) proposed that development projects absorb resources according to the cumulative Rayleigh distribution function:

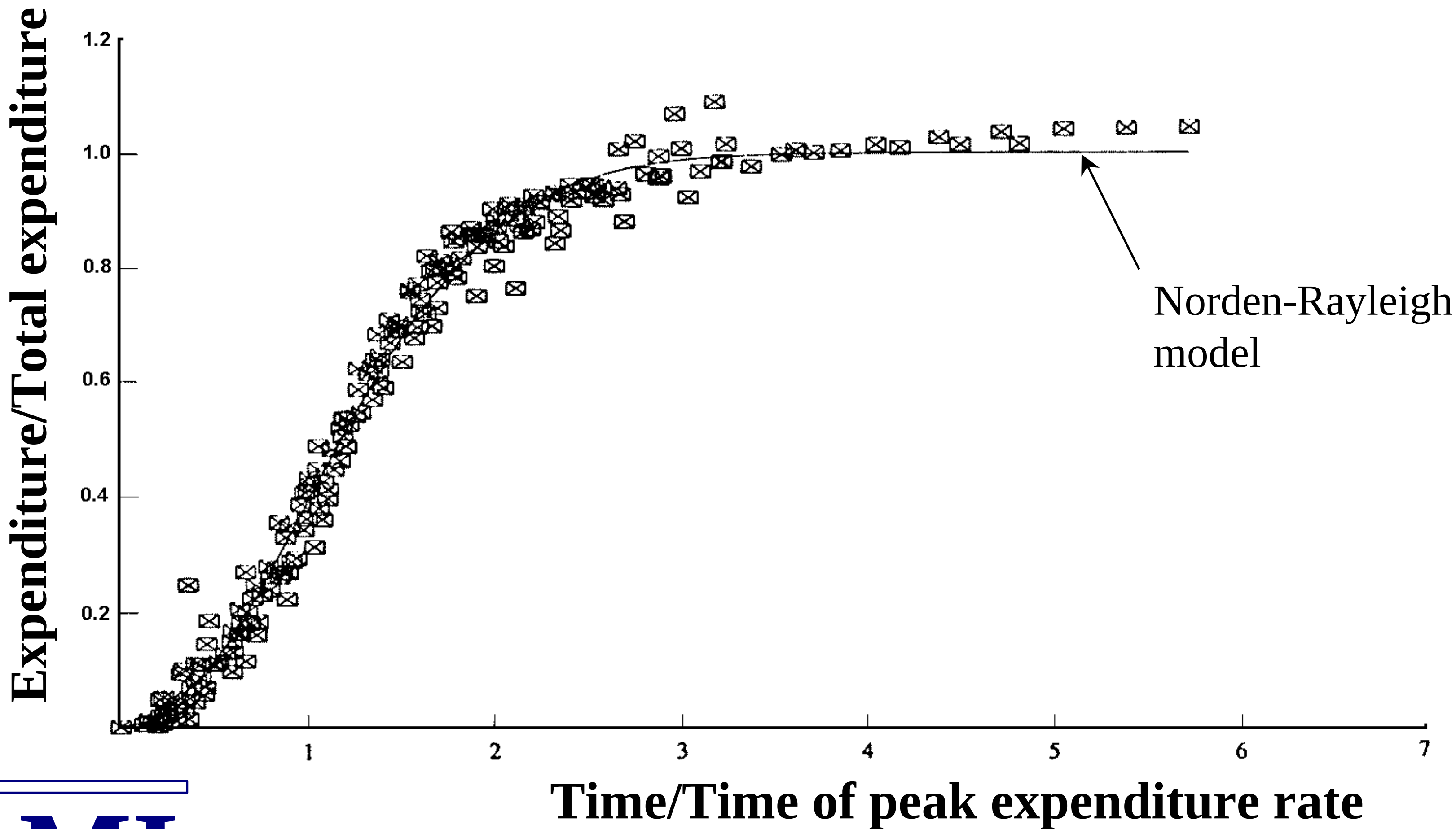
$$v(t) = d\left(1 - e^{-at^2}\right)$$

v is earned value, which may be measured by expenditures. {available, e. g., for U. S. DoD programs as Actual Cost of Work Performed (ACWP) data in Contractor Progress Reports (CPRs)}

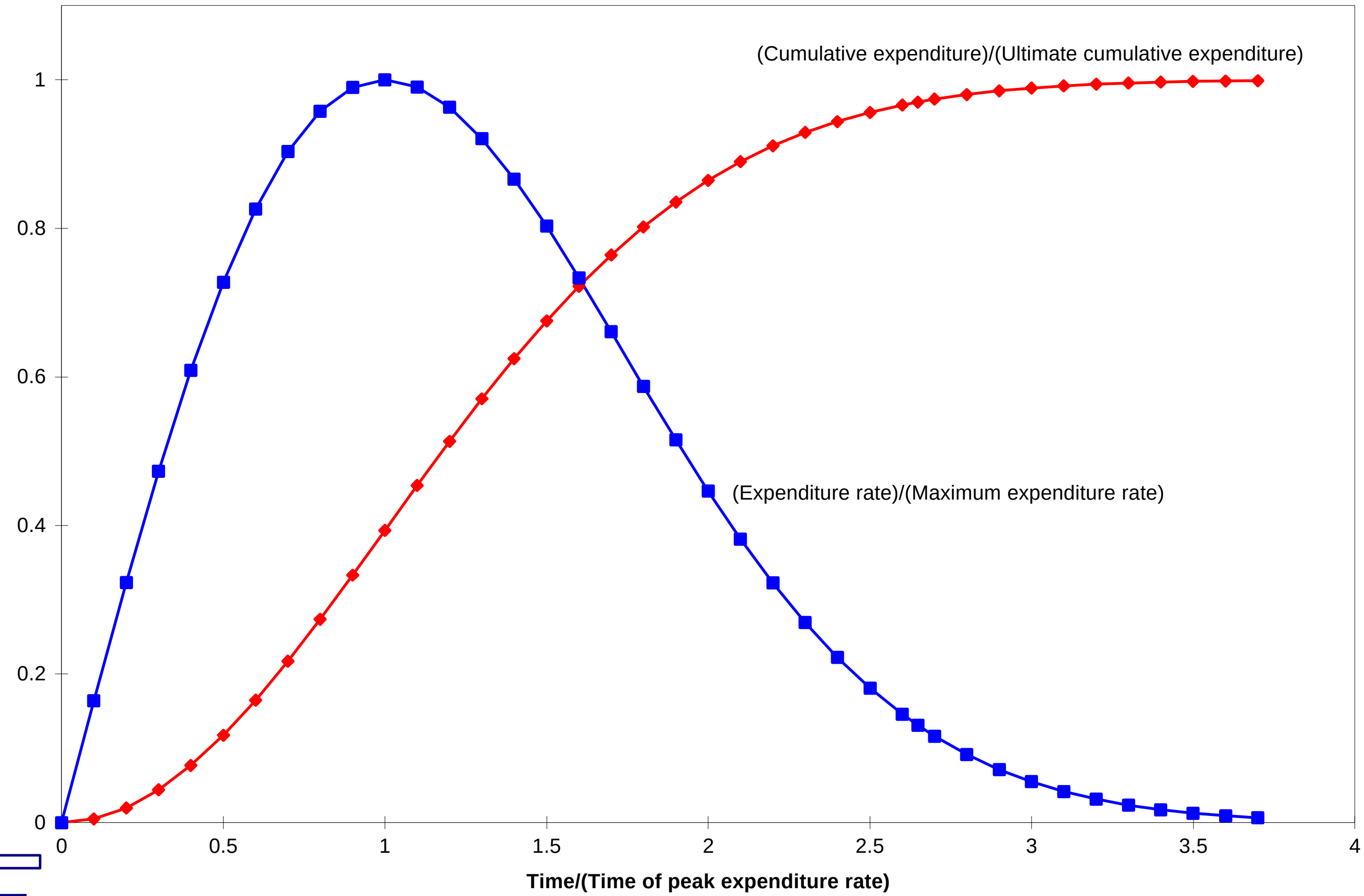
Applications of Rayleigh method

- L. Putnam, 1976 and later, to software development projects; leads to SLIM commercial estimating package
- D. Boger and students at Naval Postgraduate School, 1982 and later, to DoD development programs
- D. Lee and colleagues, OSD Cost Analysis Improvement Group, 1989 and later, to DoD development programs (Gallagher, M., and D. Lee, Mil. Op. Rsch. 2, 1996)
- G. Christle and colleagues, OUSD(A), task LMI (D. Lee and colleagues) to integrate Rayleigh analysis tool into their Contract Analysis System (CAS) (1998)

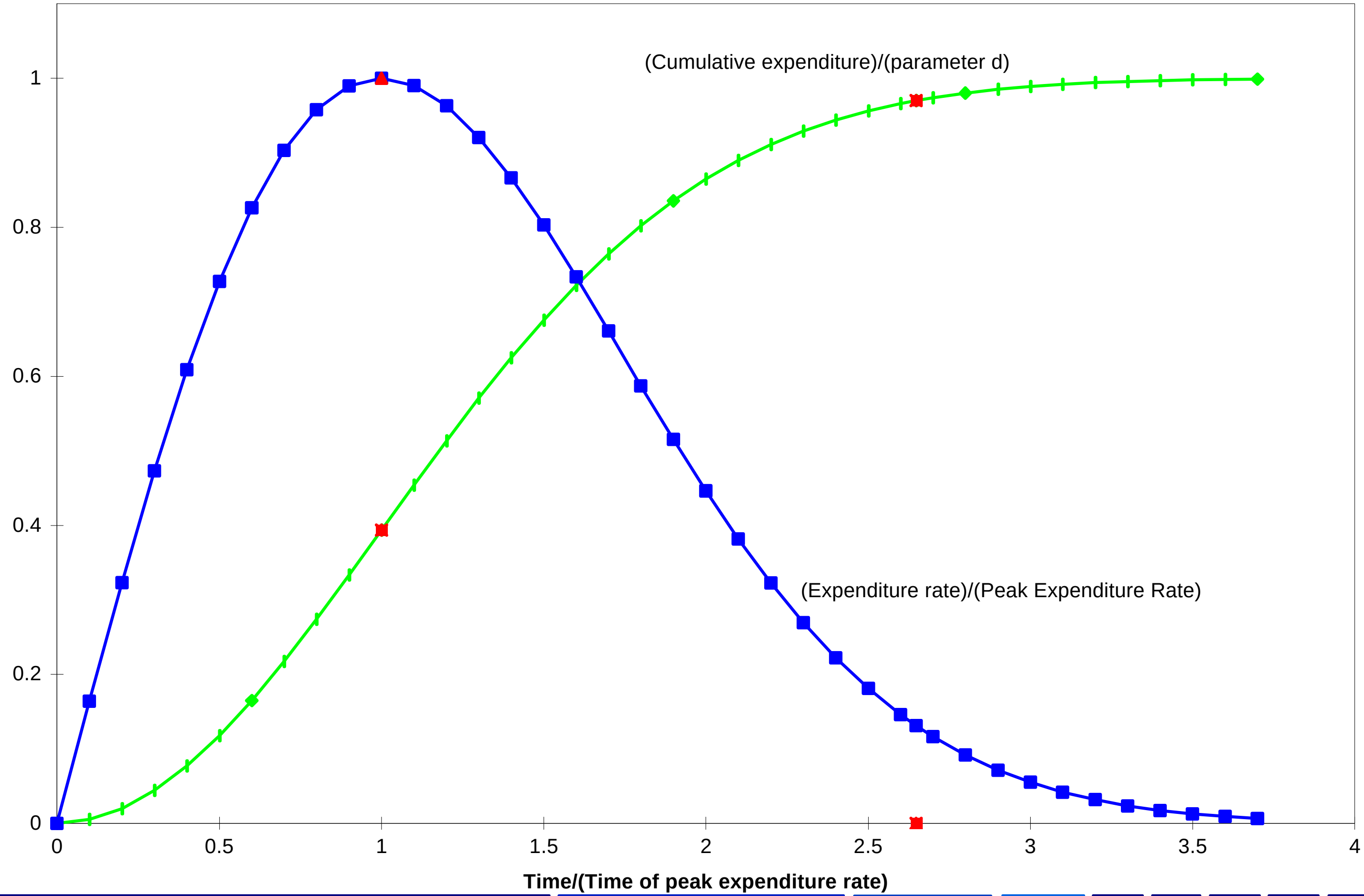
The Norden-Rayleigh model collapses data from many DoD development programs onto one curve



Shapes of Rayleigh cumulative expenditure and expenditure rate curves



Some standard points on the curves



Some standard expressions for Rayleigh curve parameters

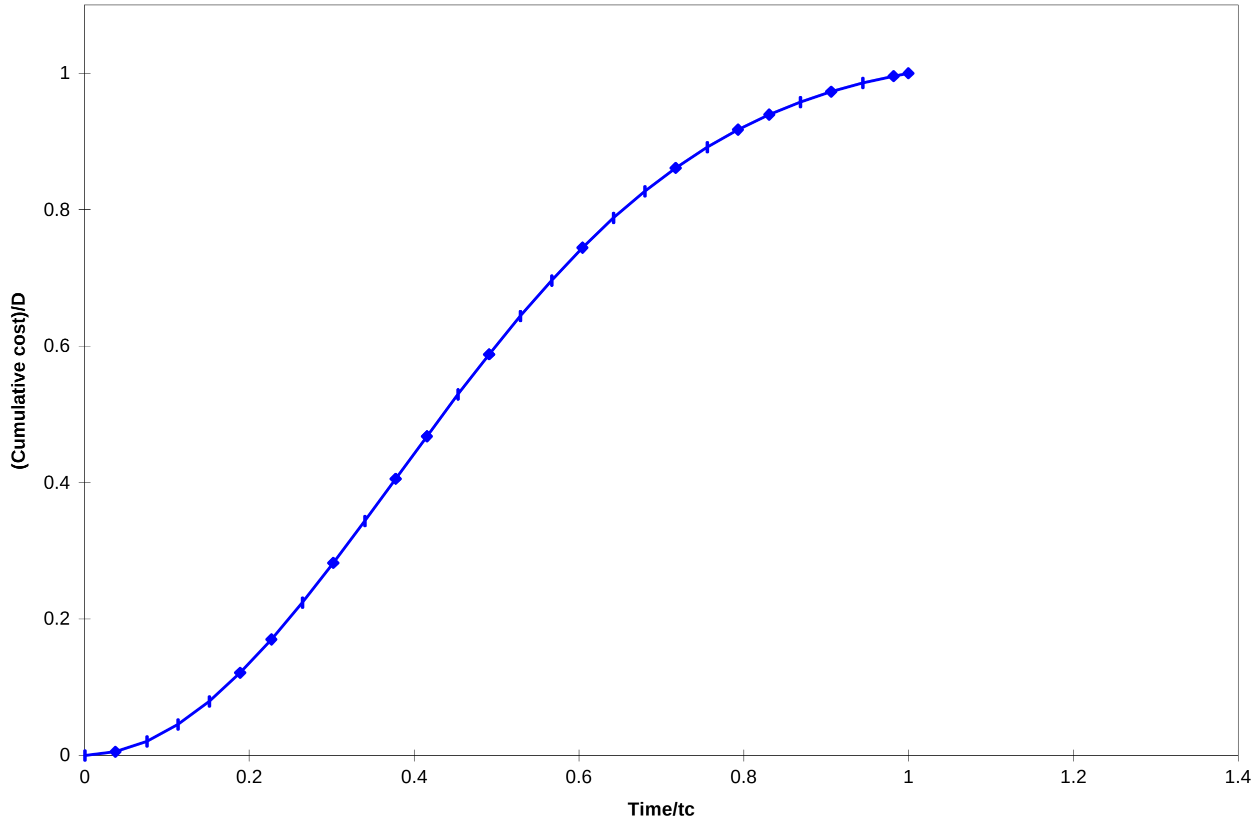
Peak expenditure time t_p and parameter a: $$t_p = \frac{1}{\sqrt{2a}} \quad ; \quad a = \frac{1}{2t_p^2}$$

By convention, finite completion time is when expenditures = 97% of parameter d:

Implies finite completion time t_c : $$t_c \equiv \sqrt{\frac{-\ln(0.03)}{a}} \quad ; \quad a = \frac{-\ln(0.03)}{t_c^2}$$

Implies final cost D = 0.97 d .

Rayleigh curve with finite completion time



Using N-R to spread a development estimate

If total estimated cost is D $BY and estimated completion time is t_c, then cumulative expenditure at time t is E $BY, where

$$E = \frac{D}{0.97}\left(1 - e^{\frac{\ln(0.03)}{t_c^2}t^2}\right)$$

Using N-R to spread a development estimate

Expenditure between times t_1 and t_2:

$$\frac{D}{0.97}\left(e^{\frac{\ln(0.03)}{t_c^2}t_1^2} - e^{\frac{\ln(0.03)}{t_c^2}t_2^2}\right) \quad \textbf{\$BY}$$

Using N-R to estimate cost-to-go and time-to-go, given initial ACWP data

- Basic idea is simple: given $(t_1, E_1), (t_2, E_2), \ldots, (t_M, E_M)$, find d and a such that $d[1 - \exp(-at^2)]$ is a "good" fit

- Then 0.97d is an estimate of total cost, and sqrt(-0.03/a) is an estimate of completion time.

Parameter estimation is computationally tractable

An example:

$$\min_{a,d} \sum_{1}^{N}\left[y_i - d\left(1 - e^{-a t_i^2}\right)\right]^2$$

Define $z_i(a) \equiv 1 - e^{-a t_i^2}$. **Then minimizing a is determined by**

$$(y \cdot z)(z \cdot z') - (y \cdot z')(z \cdot z) = 0$$

which is readily solved numerically (e. g. by bisection, or by Newton's method). Given a, d follows from

$$d = \frac{(y \cdot z)}{(z \cdot z)}$$

For devotees of linear regression:

$$y_i - y_{i-1} = d\left(e^{-at_{i-1}^2} - e^{-at_i^2}\right)$$

$$= 2adt_i^* e^{-at_i^2}(t_i - t_{i-1})$$

where t_i^***, whose existence is guaranteed by the first mean value theorem,**

is determined by $2at_i^* e^{-at_i^{*2}} = e^{-at_{i-1}^2} - e^{-at_i^2}$**. Numerical computation**

of t_i^* **is straightforward, by Newton's method or by bisection.**

For devotees of linear regression:

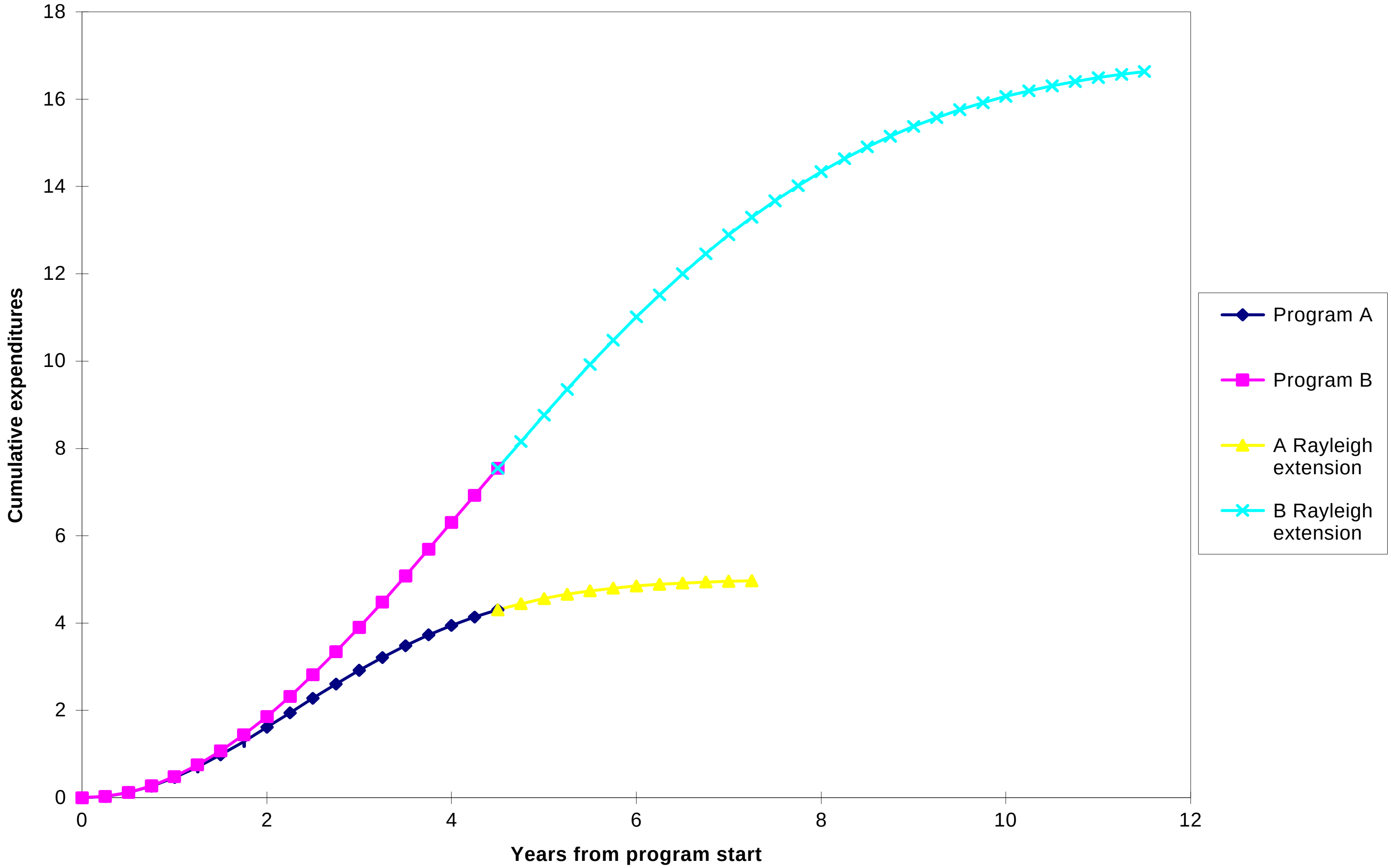
Consequently,

$$\ln\left(\frac{y_i - y_{i-1}}{t_i^*(t_i - t_{i-1})}\right) = \ln(2ad) - at_i^{*2}$$

so that one may obtain estimates for a and d from the regression coefficients

obtained by regressing $\ln\left(\frac{y_i - y_{i-1}}{t_i^*(t_i - t_{i-1})}\right)$ **on** t^{2}.

Be careful, however: the usual linear regression assumptions about the statistics of observation errors may well not be met!

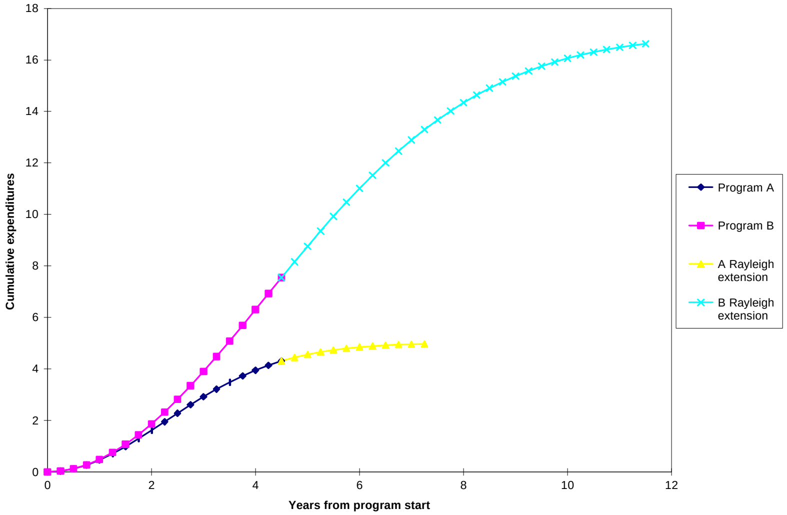


Cumulative expenditures
18
16
14
12
10
8
6
4
2
0
0
2
4
6
8
10
12
Years from program start
Program A
Program B
A Rayleigh extension
B Rayleigh extension

There's a problem if data only represent early times

- The two very different N-R curves are quite close for early times
- Problem is, that when $at^2 << 1$,

$$d\left(1 - e^{-at^2}\right) = d\left(at^2 + O(a^2t^4)\right)$$



If data come only for times small compared with t_p, it's hard to estimate both t_c and D

LMI

Resource Analysis Group

N-R for “early” data

- Difficult to estimate both total time and total cost if all data are for times less than about one-half t_p, which is about 20% of t_c.
- Often helpful to use information on *one* of completion time or total cost, to get estimates on the other that areconsistent with early cost data.
- For example, one can see if a given cost estimate is consistent with early cost data, and a given estimate of completion time.

Example N-R for "early" data

Choose a completion time t_c. Fit the model

$$v = \frac{D}{0.97}\left(1 - e^{\ln(0.03)\frac{t^2}{t_c^2}}\right)$$

to the data, by adjusting only D.

LMI

Resource Analysis Group

Using the N-R model to estimate cost-to-go and time-to-go, given ACWP data

- Apply a parameter-identification method to estimate time-scale parameter a and cost-scale parameter d, with consistent estimates of dispersion (uncertainty). Many methods are available.
- Estimate completion time and total cost, with dispersion (uncertainty) estimates, from the a and d estimates.

One method: MMAE

- Multiple Model Adaptive Estimation is a method for estimating parameters of dynamic systems, given time-history data.
- Uses set of Kalman filters, which require a parametric model for the time evolution of the system.

Gallager, M., and D. Lee, "Final-Cost Estimates for Research & Development Programs Conditioned on Realized Costs," Mil. Ops Rsch. 2, 1996, pp 51 - 65

N-R time-evolution model

If v = d[1 - exp(-at²)], then

$$\frac{dv}{dt} = 2ad\left(1 - \frac{v}{d}\right)\sqrt{-\frac{1}{a}\ln\left(1 - \frac{v}{d}\right)}$$

Evolution of earned value

If $v(t_0) = v_0$, then for $t > t_0$,

$$v = d\left[1 - e^{-a\left(t - t_0 + \sqrt{-\frac{1}{a}\ln\left(1 - \frac{v_0}{d}\right)}\right)^2}\right] \equiv V(t; a, d, t_0, v_0)$$

LMI

Kalman filter

- Given a system evolution model, Kalman filter estimates system state as a linear combination of the state predicted by the evolution model, and noisy observations of the state. For us, “state” is earned value v.

$$v(\mathrm{est}) = (1 - k)\, v(\mathrm{pred}) + k\, v(\mathrm{obs})$$

- Parameter k is called the gain of the filter

Maybeck, P., “Stochastic Models, Estimation and Control: Volume 1, Academic Press, New York, 1979

Kalman filter

$$v^{+}_{n+1}(t) = (1-k)V(t;a,d,t_{n,}v^{+}_{n}(t_{n})) + kz_{n+1}$$

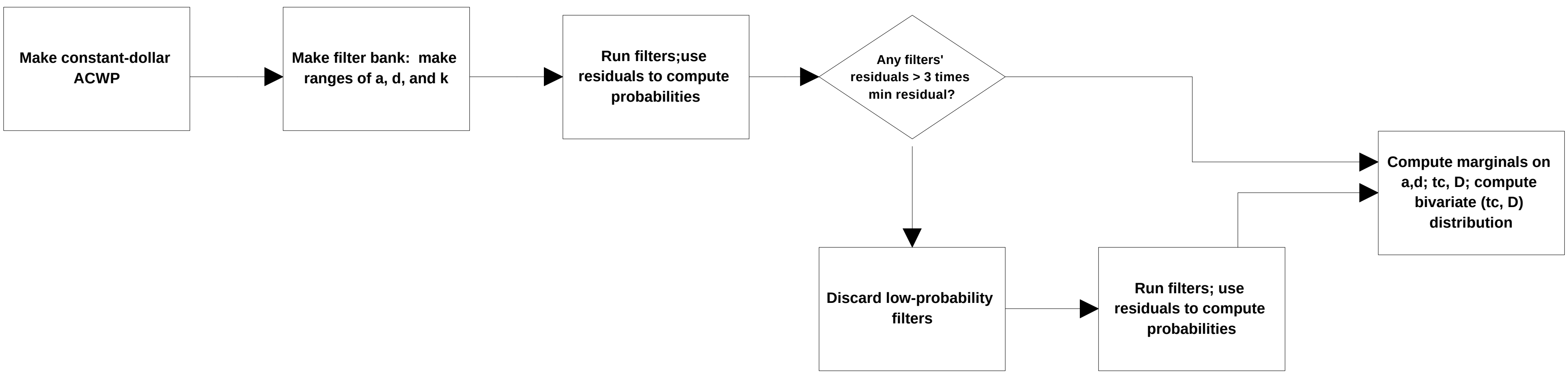
MMAE

- MMAE considers a bank of Kalman filters, each determined by three parameters (a, d, k), and determines probability that these are correct, given the ACWP data.

Maybeck, P., "Stochastic Models, Estimation, and Control: Volume 2 Academic Press, New York, 1982

Maybeck, P. S., and K. P. Hentz, "Investigation of Moving-Bank, Multiple Model Adaptive Algorithms," AIAA Journal of Guidance, Control, and Dynamics 10, 1987, pp. 771-101

Schematic of MMAE Program



Outputs from MMAE parameter identification

- Marginal distribution functions of total cost and total time, conditioned on the data
- Joint bivariate PDF of total cost and total time, conditioned on the data
- Can present costs either as $BY or as $TY

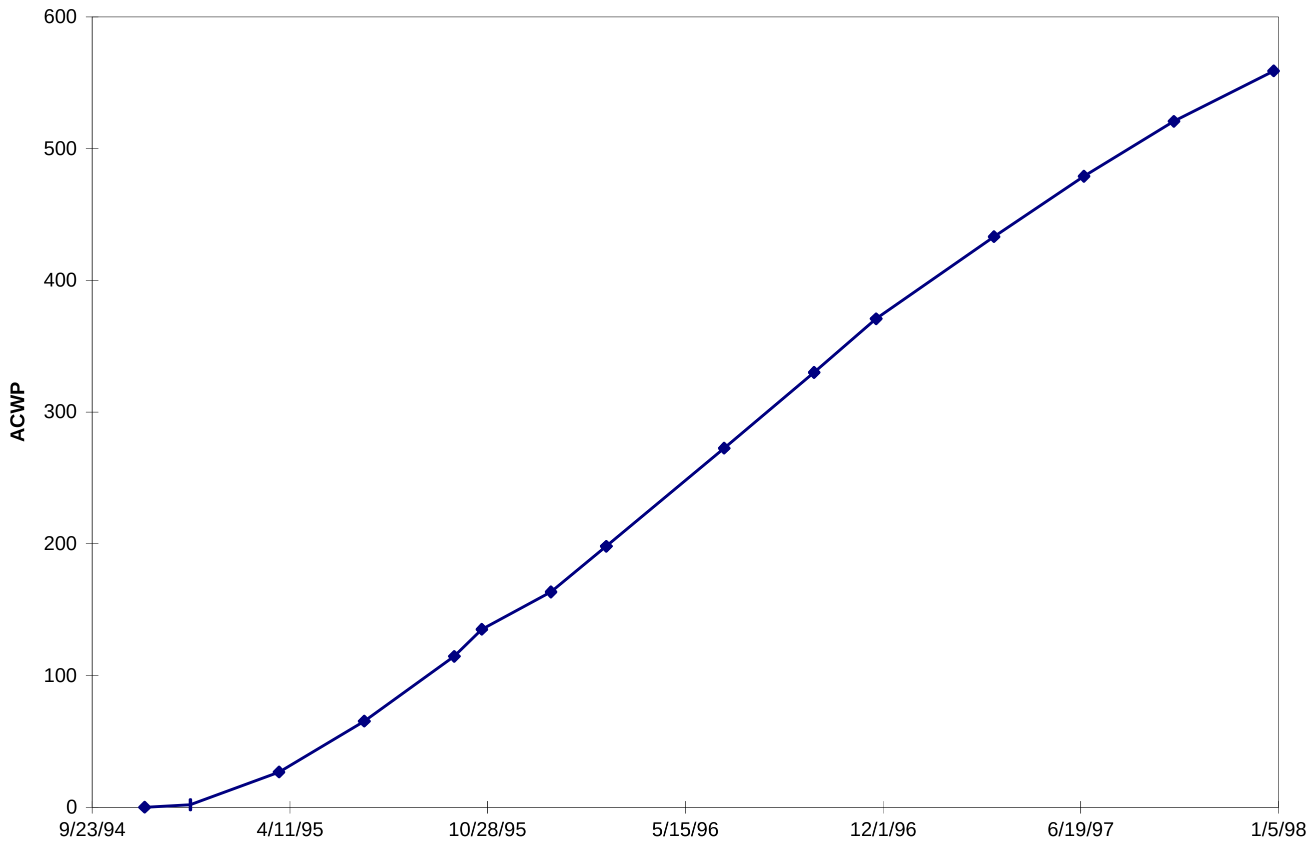
An Example

11/15/94	0
12/31/94	1.9
3/31/95	26.8
6/25/95	65.4
9/24/95	114.6
10/22/95	135.1
12/31/95	163.4
2/25/96	198.1
6/23/96	272.6
9/22/96	330
11/24/96	370.8
3/23/97	433.1
6/22/97	479
9/21/97	520.6
12/31/97	559

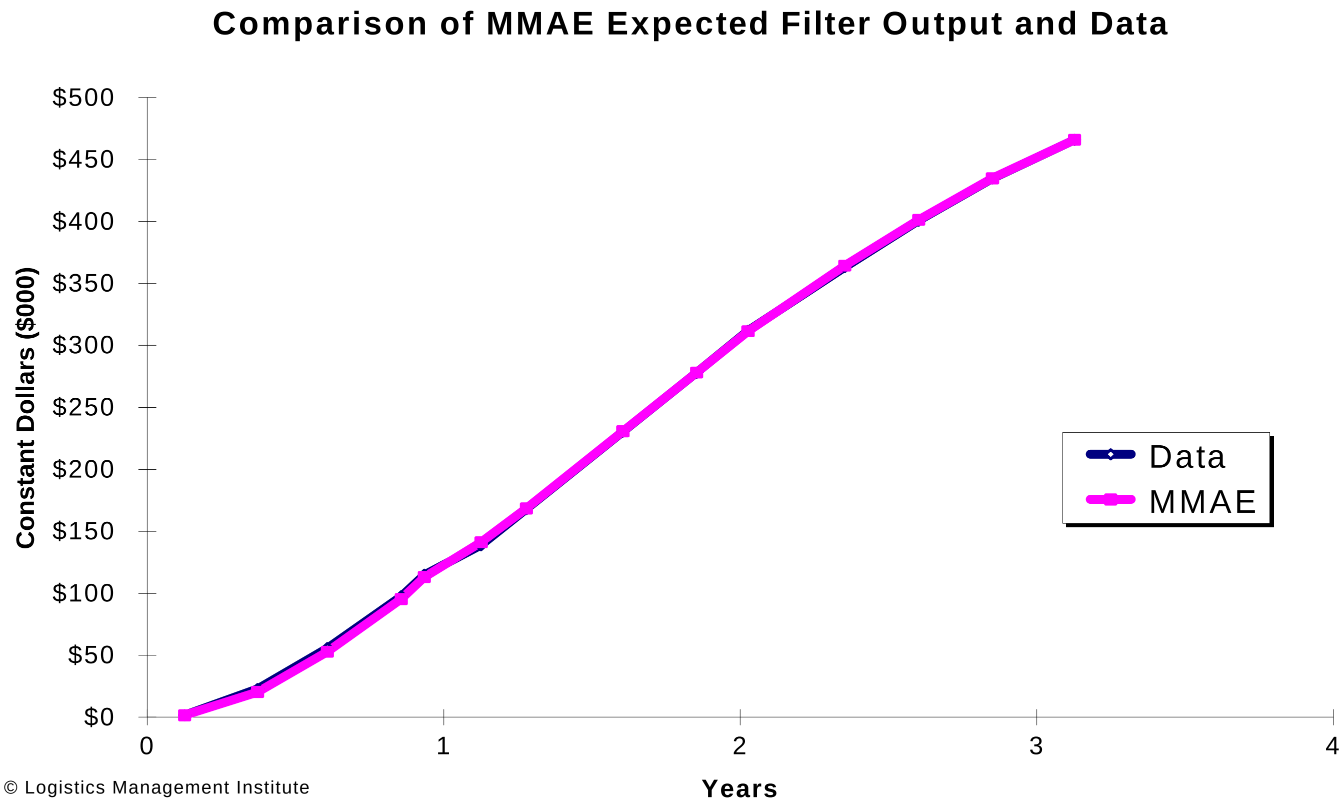
Constant-dollar ACWP

11/15/94	0				0
12/31/94	1.9	1.9	0.86011	1.634209	1.634209
3/31/95	26.8	24.9	0.85593	21.31265	22.94686
6/25/95	65.4	38.6	0.852142	32.89269	55.83954
9/24/95	114.6	49.2	0.848153	41.72912	97.56866
10/22/95	135.1	20.5	0.846929	17.36205	114.9307
12/31/95	163.4	28.3	0.843877	23.88173	138.8124
2/25/96	198.1	34.7	0.841444	29.19811	168.0106
6/23/96	272.6	74.5	0.836096	62.28919	230.2997
9/22/96	330	57.4	0.831979	47.75558	278.0553
11/24/96	370.8	40.8	0.82914	33.82891	311.8842
3/23/97	433.1	62.3	0.823804	51.32299	363.2072
6/22/97	479	45.9	0.819553	37.61749	400.8247
9/21/97	520.6	41.6	0.815318	33.91722	434.7419
12/31/97	559	38.4	0.810643	31.12867	465.8706

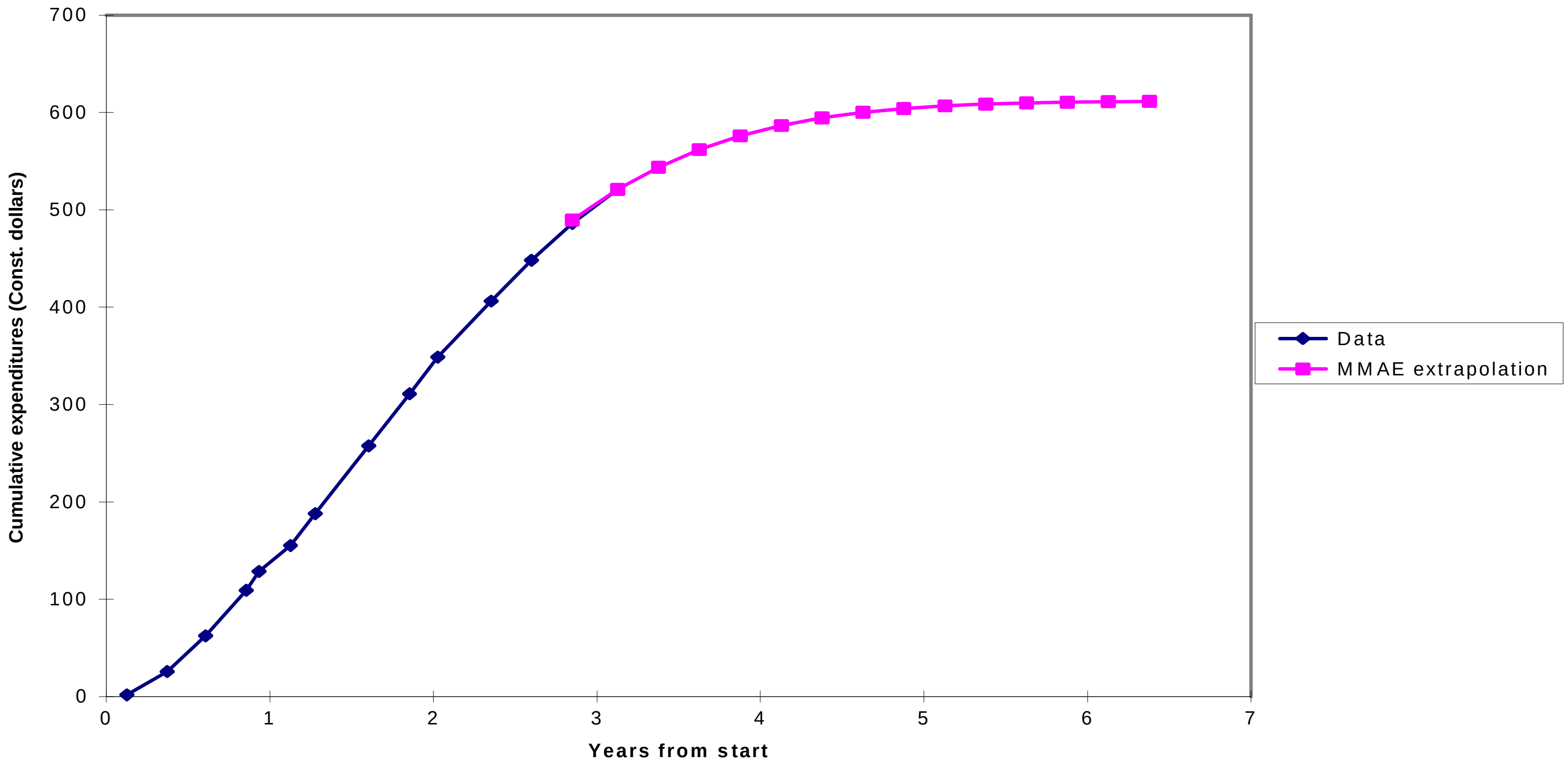
Plot of ACWP data



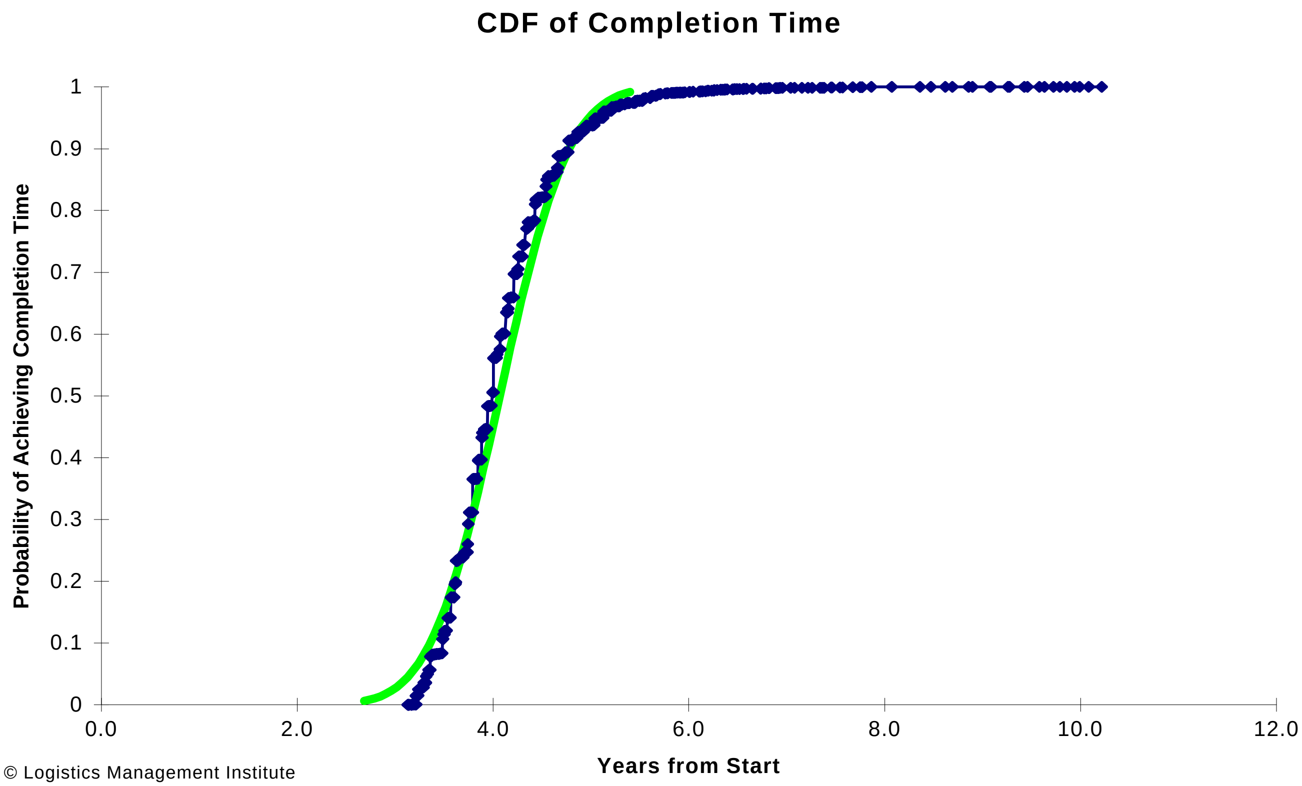
Optimal filter output and data



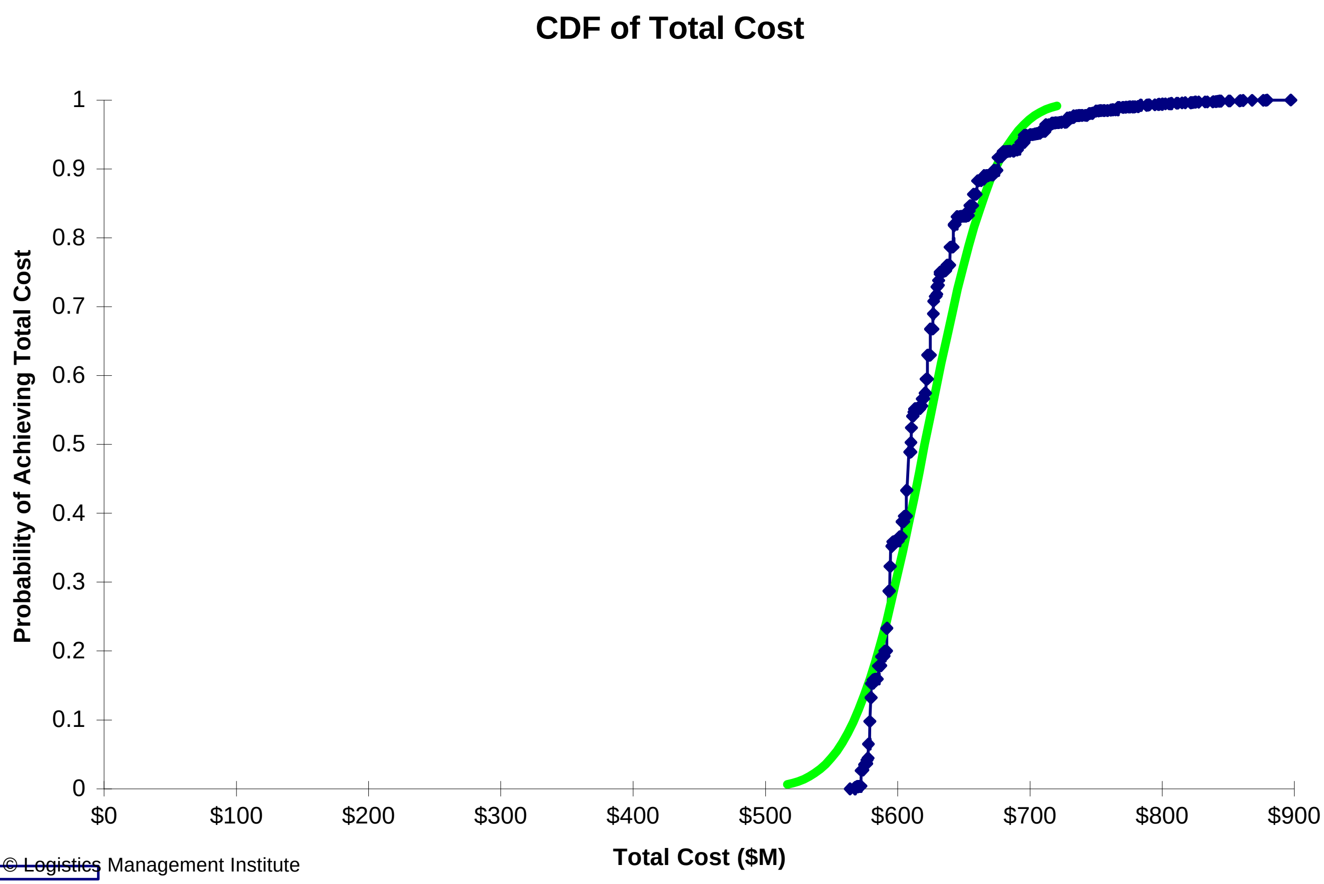
Norden-Rayleigh Extrapolation



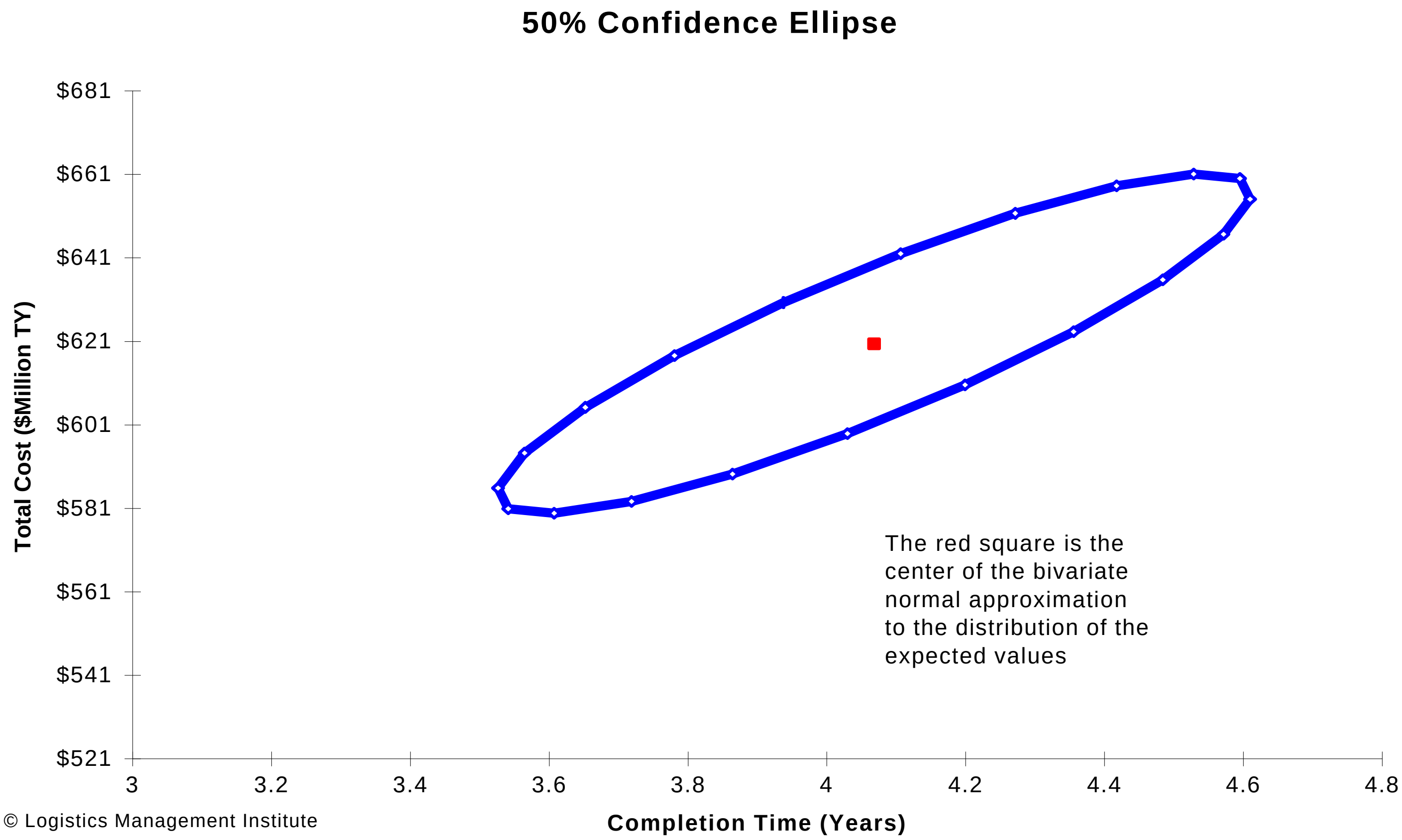
Marginal distribution of completion time



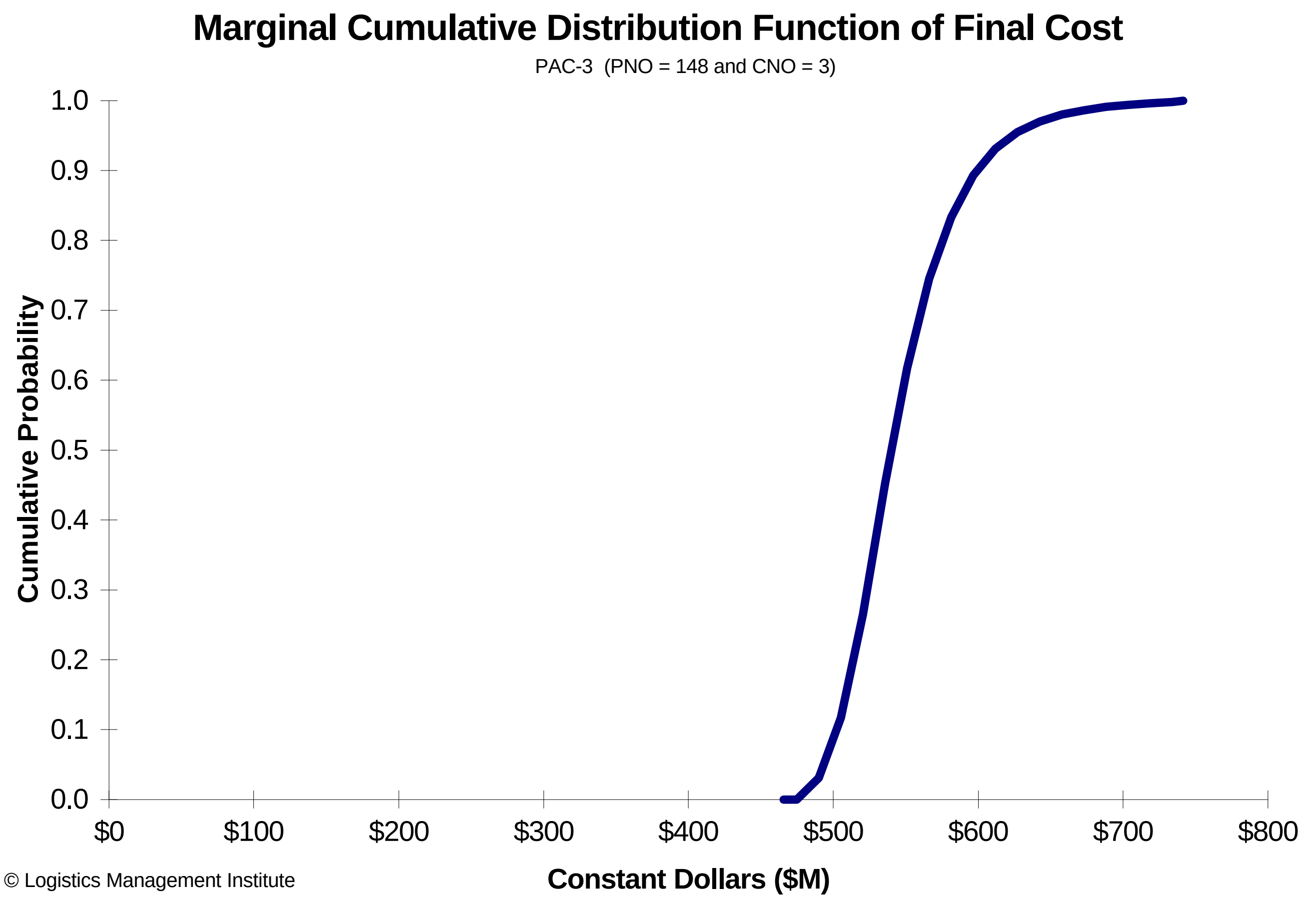
Marginal distribution of total cost



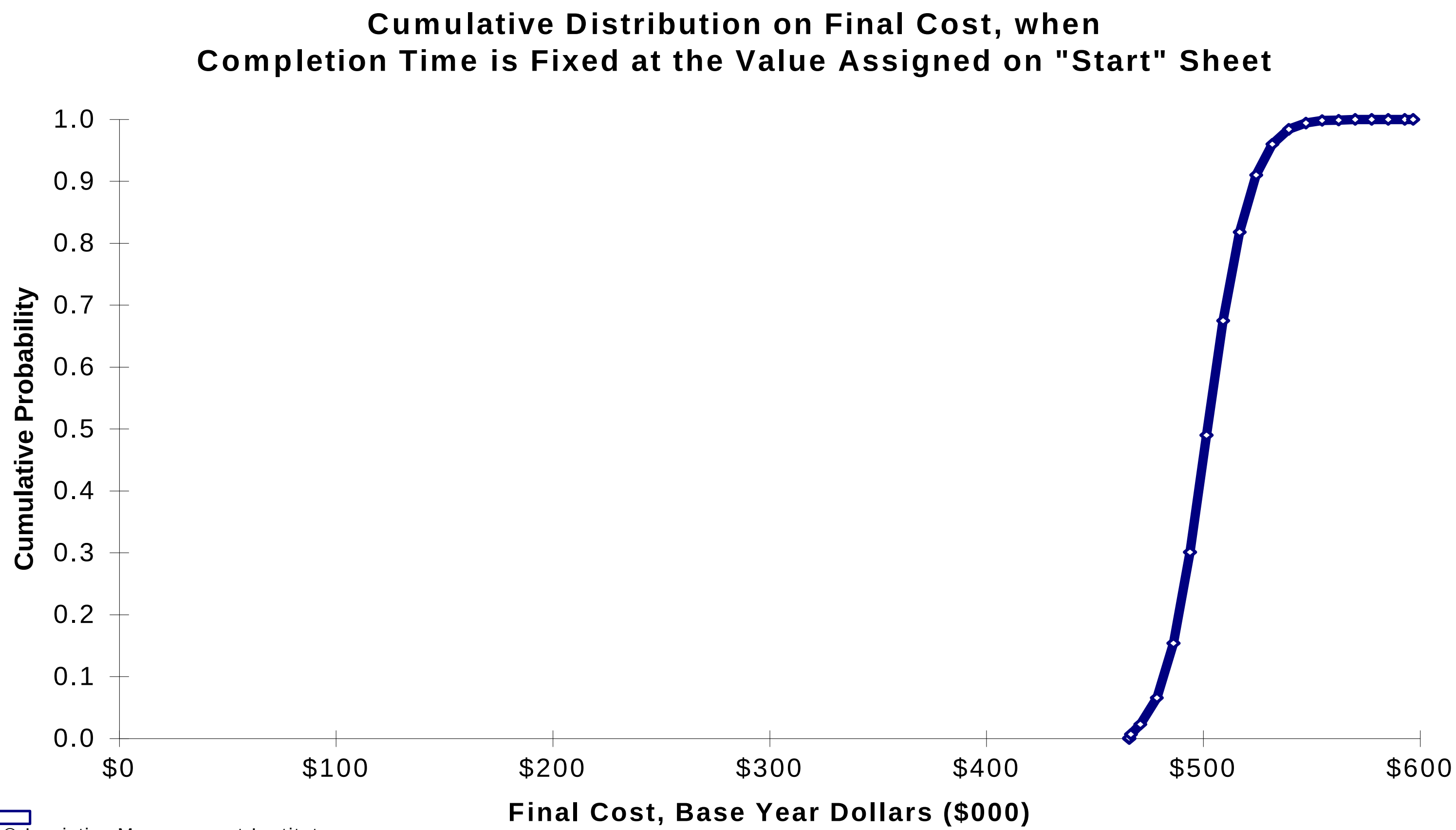
Bivariate distribution of cost and time



Cost CDF - $BY


Marginal Cumulative Distribution Function of Final Cost
PAC-3 (PNO = 148 and CNO = 3)
Cumulative Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
$0
$100
$200
$300
$400
$500
$600
$700
$800
Constant Dollars ($M)

Cost CDF for 4-year Program



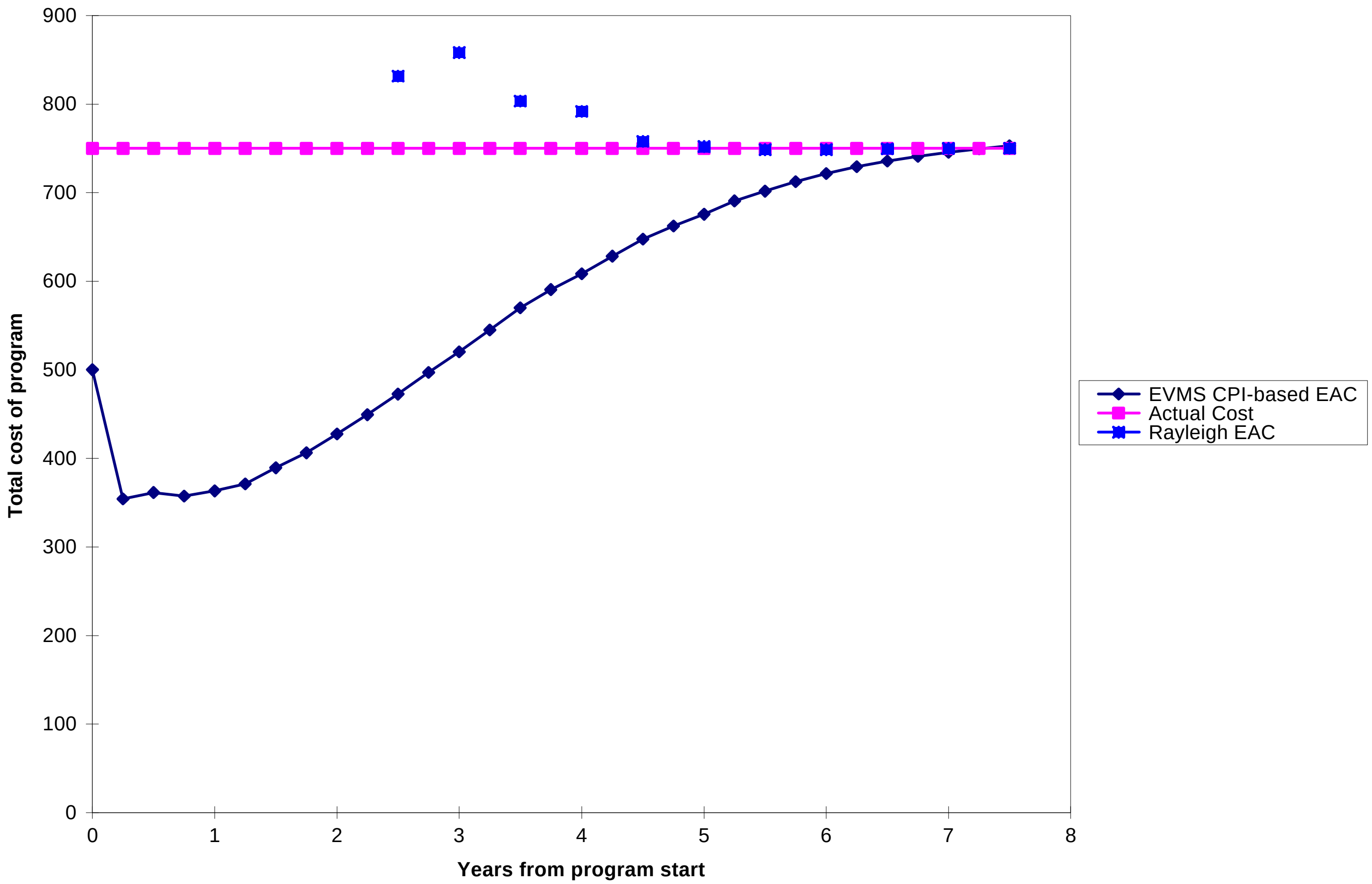
LMI

Resource Analysis Group

Comparison with EVMS (formerly C/SCSC)

- If EVMS planning estimates include all required work, and if they under- or over-estimate by exactly the same ratio for all parts of the project, EVMS EAC based on CPI will be accurate
- If planning estimates are better for initial phases than for later ones, initial EVMS forecasts will be optimistic.

Example



Total cost of program
900
800
700
600
500
400
300
200
100
0
0
1
2
3
4
5
6
7
8
Years from program start
EVMS CPI-based EAC
Actual Cost
Rayleigh EAC

Rayleigh isn't the only possible expenditure vs time function

- Basic ingredient is time evolution model

$$\frac{dv}{dt} = F(v)$$

which may be identified using non-parametric methods of system identification

Summary

- Rayleigh analysis gives parametric model of development program expenditures
- Method strongly supported by data from U. S. DoD development programs
- Generates forecasts cost-to-go and time-to-go; time-phasing of total-cost estimates